Volume 12 • Number 2 • 2009
VALUE IN HEALTH

What Outcomes Are Important to Patients with Long Term Conditions? A Discrete Choice Experiment

Gerry Richardson, PhD,[1] Chris Bojke, MSc,[2] Anne Kennedy, PhD,[3] David Reeves, PhD,[3] Peter Bower, PhD,[3] Victoria Lee, PhD,[3] Elizabeth Middleton, PhD,[3] Caroline Gardner, BSc,[3] Claire Gately, MSc,[3] Anne Rogers, PhD[3]

[1]University of York, Heslington, York, UK; [2]PharMerit, York, UK; [3]University of Manchester, Manchester, UK

ABSTRACT

Objective: To assess how much patients with long-term conditions value self-efficacy (i.e., confidence in their ability to manage their condition) compared with other health outcomes, including measures of quality of life, and process outcomes including access to General Practitioners.

Methods: Discrete Choice Experiment (DCE) set in UK community settings. Participants: 367 patients (mean age 57.5) living in the community with a wide range of self-defined long-term conditions. Main outcome measures: the relative value that individuals place on four specific outcomes, namely, self-efficacy, Health Related Quality of Life (HRQoL), access to General Practitioners, and level of isolation.

Results: Most responders completed their questionnaire in a consistent manner. Most valuations of outcomes were in the expected direction and were statistically significant. A substantial minority of responders exhibited counter-intuitive preferences. The existence of a significant constant in all models raised concerns about model misspecification. Nevertheless, all models showed that participants were willing to trade substantial reductions in their HRQoL for improvements in their self-efficacy.

Conclusions: The majority of patients with chronic conditions were able to complete the DCE questionnaires. However, the existence of counter-intuitive preferences and evidence of model misspecification require further investigation. These issues are largely overlooked in the health economics literature. Self-efficacy is an important outcome for this group and is not included explicitly in conventional HRQoL measures. This is potentially important where decisions are made on the basis of cost-effectiveness using Quality Adjusted Life Years as the metric. Exclusion of these outcomes may lead to the cost-effectiveness of these interventions being understated.

Keywords: Discrete Choice Experiment, health economics, Quality Adjusted Life-Years, quality of life.

Introduction

Recent policy has targeted self-care as a means to improve patient outcomes and reduce costs [1,2]). Self-care has been defined as care taken by individuals toward their own health and well-being [2]. Interventions have been designed that support individuals' ability to self-care, for example, the Expert Patient Programme (EPP) based on the chronic disease self-management program developed in the United States by Lorig [3,4]. The "EPP" aims to provide self-care support to any individual with a chronic condition in England and is a generic, lay led, group program involving six-weekly sessions. The EPP is designed to enable participants to develop appropriate self-care skills including patient's "self-efficacy" [5]. The concept of self-efficacy refers to a psychological state which relates to an individual's confidence that they can achieve some task (such as managing the symptoms of their disorder or engage in regular exercise) [6–8], and is one of the most important concepts in modern psychological approaches to understanding health behavior [9]. Self-efficacy is increasingly accepted as a mediating variable and an outcome of self-management programs but this has become normative without evidence of what patients most value in managing long-term conditions.

A recent trial [5] demonstrated that a UK version of the chronic disease self-management program was effective at improving self-efficacy, and a cost-effectiveness analysis based on the same clinical trial [10] generated Quality Adjusted Life-Year (QALY) gains for these interventions using the EQ5D instrument. The EQ5D instrument measures Health Related Quality of Life (HRQoL), across five dimensions, namely mobility, ability to self-care, ability to perform usual activities, level of pain/discomfort, and level of anxiety/depression. The cost-effectiveness analysis concluded that the EPP intervention was likely to provide a cost-effective alternative to usual care in people with long-term conditions at commonly used threshold values of a QALY.

Thus QALYs, often generated from EQ5D, have a commonly expressed value [11] but they may not incorporate all the outcomes that are of interest. In contrast, while self-efficacy is undoubtedly an outcome of interest (at least for practitioners and researchers), we have no knowledge of whether self-efficacy is "of value" per se, or indeed what that value might be. This leads to problems of interpretation as decision-makers cannot assess the relative merits of self-efficacy compared with HRQoL. In the example above, the EPP was deemed to be likely to be cost-effective based solely on the QALY, although there was a large degree of uncertainty around the decision. Valuing self-efficacy (or other relevant outcomes) in terms of QALYs gives decision-makers additional information and can be incorporated into the cost-effectiveness analysis and could reduce the uncertainty around a decision.

Discrete Choice experiments (DCEs) are one method of either expanding the measure of outcome or incorporating factors other than health outcomes [12]. DCEs are based on the premise that the benefits associated with health-care interventions can be expressed in terms of the "attributes or characteristics" of that intervention [13] and the "attributes or characteristics" of the person valuing them [14,15].

Process outcomes may be important in the evaluation of health-care technologies, and DCE enables the relative values of these outcomes to be assessed [12]. For example, speed of access to health care, who provides that health care and where it is provided, are often considered as important aspects of health care that may not be captured by a measure of HRQoL [12].

Address correspondence to: Gerry Richardson, Centre for Health Economics, Alcuin "A" Block, University of York, Heslington, York YO10 5DD, UK. E-mail: gar2@york.ac.uk
10.1111/j.1524-4733.2008.00419.x

 1098-3015/09/331 331–339

DCEs have been used frequently in the health economics literature to estimate preferences in miscarriage management [16], management of prostate cancer [17], as well as in a variety of other conditions and in numerous settings [18–29]. However, DCEs have not been previously employed to investigate the trade-off between HRQoL, psychological outcome measures such as self-efficacy, and social outcomes such as isolation and/or process outcomes such as General Practitioner (GP) access. One advantage of the DCE methodology is that it enables the individual to simultaneously compare and value a number of different attributes. This may have the added benefit of limiting the "salience" problem identified with standard gamble techniques, where there is discordance between stated preferences and actual choices because of the experiment focusing the individuals' concentration on one specific attribute [30], although it is acknowledged that responders may still not focus on the true opportunity cost of their choice [31].

This article describes a DCE conducted to examine the relationship between HRQoL and other outcomes which may be of relevance to patients with long-term conditions. In addition, the estimation of rates of substitution between the QALY and self-efficacy enables decision-makers to include these outcomes in their assessment of cost-effectiveness.

Methods

The authors conducted a DCE: a questionnaire based stated choice method, to explore the outcomes that are most valued by patients. There are several recognized steps for conducting a DCE [32,33], described below.

Defining Attributes/Outcomes

Qualitative interviews and focus groups carried out alongside a Randomised Controlled Trial of the EPP identified isolation and access to health professionals as a major influence on individual well-being [34]. In addition a systematic review of the published economic evaluation literature identified self-efficacy as an important outcome that patients valued [35]. The attributes and their levels were chosen to be plausible for the responders to answer [32].

Based on these sources, three outcomes were selected for inclusion into the study and access to GPs, level of social isolation and level of self-efficacy (patients' confidence in their ability to manage their condition). In addition, because the rationale of the study was to assess the importance of self-efficacy and the other outcomes relative to HRQoL and examine whether such outcomes could be included in cost-effectiveness analysis, a measure of HRQoL was included. As it is frequently used in economic evaluation and has previously been used in DCEs [36,37], the EQ5D was used as the basis for measuring HRQoL. EQ5D measures patient health status across five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with three possible responses (no problems, moderate problems, or severe problems) for each dimension. This locates each participant into one of 243 mutually exclusive health states. Clearly, including this number of levels of an attribute is impractical. Three states were selected as levels for this attribute to maintain the statistical efficiency of the experiment. The three states showed a clear ranking between them and were all levels that were considered plausible for this patient group [32] and that they were likely to have experienced.

The use of "psychological" outcomes in DCEs, such as self-efficacy, is potentially limited by the interpretation responders place on them. It could be argued that these outcomes depend on emotional state of the individual or could be merely acting as surrogates for other outcomes (such as HRQoL). However, evidence from the pilot study suggests that this may not be a major problem in this study. It should also be noted that these problems are not unique to this DCE, or DCEs in general. Indeed, similar issues of interpretation have been identified with "objective" measures of health status such as the SF36 [38,39]. The generalized Linear Latent and mixed models (GLLAMM) model described below also allows valuations of attributes from different groups with systematically different values. This model estimates that the proportion of individuals in each "latent class," and allows the preferences of these different classes to be estimated.

The endogeneity of this measure is also a potential problem. Ideally, it may be better to describe objective outcomes (such as a health state) and derive values for this health state from individuals rather than asking them to value a subjective health state such as confidence. While this is a limitation of the study, the policy aim was to establish the value of self-efficacy (or isolation or GP access) relative to HRQoL. This is an important policy question as the EPP was purchased and rolled out based on the results of experiments showing improvements in self-efficacy. Self-efficacy could be considered an important outcome in its own right, not just as a surrogate for future improved health.

To improve the statistical efficiency of the experiment (that is to enhance the power of the study to detect the impact of moving between levels of attributes), the number of levels in each attribute should be multiples of each other [40]. That is if we have four attributes, we could have three with four levels and one with two levels, or even one with eight levels. We should not, however, have some attributes with two levels and others with three. The practical implication for this DCE was that there should be four attributes each with three levels (as having three levels for each clearly satisfies the above requirement). Having more than three levels for each attribute increases the number of questions responders is asked and therefore increases the cognitive burden considerably.

Not only is this design statistically efficient, but it also enables some two-way interactions terms to be examined in a longer version of the questionnaire described below. These interaction terms are typically ignored in DCEs in the health economics field.

The attributes and levels identified from this process are presented in Table 1.

Short and Long Questionnaires

Two questionnaires were developed. Both questionnaires used the same attributes with identical levels; however, the number of questions differed. The shorter questionnaire contained 10 questions (8 unique questions and 1 repeated question), while the longer questionnaire contained 28 questions (26 unique and 1 repeated). This allowed us to test whether questionnaire length (and therefore cognitive burden) impacted on response rate. In addition, the longer questionnaire enables the inclusion of interaction terms in the model. Interaction effects are ignored in most DCEs in the health economics literature as they add to complexity and to the cognitive burden for respondents. In this DCE the possibility that there were "interactions" between the attributes described in the DCE was explored. For example, being in a bad health state may reduce an individuals' utility by 0.2 and having low levels of confidence may reduce utility reduces by 0.1. However, having both bad health and low confidence may result in a drop of utility greater than 0.3.

The main results section presents results where the responses to short and long questionnaires were pooled, so that both

Table 1 Attributes/outcomes for Discrete Choice Experiment

Attribute	Levels
Health Related Quality of Life	1. No problems with mobility, usual activities, self-care or anxiety/depression. Moderate pain/discomfort 2. No problems with usual activities, self-care or anxiety/depression. Some problems with mobility and moderate pain/discomfort 3. No problems with usual activities. Some problems with mobility and self-care. Also moderate pain/discomfort and moderate anxiety/depression
Level of confidence	1. Totally confident in ability to manage condition 2. Moderately confident in ability to manage condition 3. Not at all confident in ability to manage condition
Access to General Practitioner	1. General Practitioner appointment tomorrow 2. General Practitioner appointment in one week 3. General Practitioner appointment in 3 weeks
Level of isolation	1. See friends/relatives daily 2. See friends/relatives every few days 3. See friends/relatives rarely

questionnaires have an impact on the coefficients in the "main effect" model. Separate analyses were performed to identify any differences between questionnaires, for example, there may be more inconsistent responses on the longer questionnaire.

Questionnaire Methods

Hypothetical choice sets were created using a fractional factorial design with fold over [41]. For this DCE, the first scenario was generated from a catalogue of designs [41] and had health state at level 1, self-efficacy at level 2, GP access at level 2, and isolation at level 3. This can be represented as 1223 and would form one scenario in the choice set. To generate the comparison, which responders will choose between, levels are "folded over." Thus, in this DCE, level 1 becomes level 2, 2 becomes 3 and 3 becomes 1. Therefore, 1223 becomes 2331 when folded over. The comparator then becomes health state at level 2, self-efficacy at level 3, GP access at level 3, and isolation at level 1. This creates one choice set. The procedure is repeated for the number of questions to be asked to create the full questionnaire.

This design is commonly used in DCEs in health and other disciplines, as it allows for a relatively efficient estimation of the coefficients around attributes [40]. These designs inevitably lose some information when compared with the full factorial design [40] (where the full range of combinations of attributes and their levels are presented), but reduces the cognitive burden for respondents. Each choice set required the respondent to select one of the unlabelled options A or B (i.e., it was a "forced choice"). Some authors [42] have suggested that additional options should be included (such as an opt-out clause) so that the experiment is not a forced choice, while others [43] have claimed that this may increase the number of neutral responses and thereby the "power" of the study. Pilot work on this study indicated that neutral responses were likely in this DCE and a forced choice was chosen as appropriate.

The design of questionnaires was orthogonal, that is main effects could be estimated without correlation with other main effects or interactions [41,44] (i.e., the individual attributes have an independent effect on utility). While this is usually a desirable characteristic, it does not allow for any a priori information about responders preferences [45]. In this instance however, although we had identified important attributes, we had no a priori knowledge of their relative importance and an orthogonal design was appropriate. Other desirable characteristics [36] of DCE's are level balance, where each level of each attribute appears the same amount of times in the scenarios, and minimal overlap, where the attribute level is the same in a given choice set should be minimized (thus for example, if Option A has health state 1, Option B should ideally not have health state at level 1).

This DCE is an unlabelled experiment in that the labels (A and B) give no additional information. The only information individuals have is provided by the attributes and their levels [40]. There is no hypothesized good or service associated with the attributes and their levels. An example of the questions facing patients is presented in Figure 1.

A		*B*
• You have no problems walking about no problems with self-care no problems with usual activities **moderate** pain or discomfort no anxiety or depression • You are **not** confident you can manage your condition • You can have a GP appointment **tomorrow** • You see your friends or relatives **daily**	**OR**	• You have **some** problems walking about no problems with self-care no problems with usual activities **moderate** pain or discomfort no anxiety or depression • You are **totally** confident you can manage your condition • You can have a GP appointment in **3 days'** time • You **rarely** see friends or relatives

Please tick one box:

Choice A ☐ **Choice B** ☐

Figure 1 Imagine that you can have either Option A or Option B, which would you choose? If you would choose the option where you have moderate pain or discomfort, are not confident that you can manage your condition but you can have a General Practitioner (GP) appointment tomorrow and you see friends or relatives daily (i.e., everything in column A) then choose Option A. However, if you would prefer the option where you have moderate pain or discomfort as well as having some problems with walking, but you are totally confident that you can manage your condition with a GP appointment in 3 days' time, but you rarely see friends or relatives (i.e., everything in column B), then choose Option B.

Consistency

One question was repeated to check that patients responded in the same way to each, as a check for "rationality" [46]. Specifically, this repeated question tests for consistency in response. Where the responder gave different answers to the same question, this was considered "irrational," for the purposes of this analysis. There is a large literature on testing for rationality of responses in DCEs (see for example [47–52]), and a degree of controversy over what constitutes "irrationality," how to test for rationality and how to proceed with "irrational" responses.

As there is a lack of consensus on testing for, and the appropriate method for dealing with, "irrational" responses [47,51,52], analyses including and excluding these "irrational" responses were conducted.

Pilot Study

Questionnaires were tested via a pilot study. Issues of interest included whether patients understood the questionnaire, whether the attributes were traded, whether one attribute dominated the others, and whether the responses were internally consistent. For the pilot study, 27 individuals with a chronic health problem completed a questionnaire and were then telephone interviewed. In general, patients understood the exercise, although the hypothetical nature of the experiment was not emphasized sufficiently. The hypothetical nature of the study was stressed more for the main study questionnaire. Several responders felt that the long questionnaire was repetitive, but most patients agreed with the choice of attributes. Some responders identified other attributes which they felt may be important such as continuity of care and the quality of the interaction with health professional(s). Several responders stated that they would have opted out of the choice where there was little perceived difference between the options offered, if they had been given an opt-out option. These respondents were ultimately able to make a choice between the options, and for this reason the opt-out was excluded from this questionnaire. The downside to the omission of the opt-out is that there is a possibility of overstating the importance or relative weight of attributes in the DCE.

Responders were asked specifically whether they felt the attributes measured in the study were important. Several stated that confidence was very important, as it gave them "permission to manage their condition;" others felt that confidence was not an issue for them and that, for example, "pain was the only issue for me." While these are far from conclusive, this would appear to support the notion that responders were treating the confidence dimension separately from the quality of life dimension.

Main Study

Postal questionnaires were sent to a sample of individuals who had taken part in the EPP trial. Patients, who were involved in the randomized controlled trial (RCT) of the EPP and had not indicated that they would prefer not to receive any more questionnaires, were included in the study. Thus the study sample was patients with a (self-defined) chronic condition and there were no exclusion criteria. A free phone number was provided for any questions patients might have had. Patients who did not respond after 2 weeks received a written reminder.

Model Estimation

The DCE was analyzed by treating each choice between pairs as a single observation.

In each choice the rational respondent will choose the option that yields the highest level of utility to them and so in choice j individual i will choose B over A if $U_{jBi} > U_{jAi}$. From the construction of these indirect utility functions it can be shown that the difference in utility is due to the difference in attributes between Option A and B in choice j and the error term.

$$\begin{aligned} y^*_{ij} = U_{jBi\text{-}jAi} = U_{jBi} - U_{jAi} = (\beta_0 + \beta_1 x_{jB} + \ldots + \beta_N x_{NjB} + \\ \gamma_0 + \gamma_1 s_{1i} + \ldots + \gamma_M s_{Mi} + \varepsilon_{jBi}) - (\beta_0 + \beta_1 x_{jA} + \ldots + \\ \beta_N x_{NjA} + \gamma_0 + \gamma_1 s_{1i} + \ldots + \gamma_M s_{Mi} + \varepsilon_{jAi}) = \beta_1 \Delta x_1 + \\ \ldots + \beta_N \Delta x_N + \varepsilon_{jBi} - \varepsilon_{jAi} = \beta_1 x_{1j} + \ldots + \beta_N x_{Nj} + \varepsilon_{ji} \end{aligned}$$

where $x_{1j} = x_{1jB} - x_{1jA}$ and $\varepsilon_{ji} = (\varepsilon_{jBi} - \varepsilon_{jAi})$

As specification individual characteristics appear in equal measure on both sides of the difference equation, they simply difference out and hence the decision to choose A over B or B over A is independent of individuals' observable characteristics, including their health status. In this homogenous model, individuals may value the profiles differently, but the expectation of the difference between the utilities of the two profiles will be the same for all individuals, but will differ in reality subject to the realization of the error terms.

The dependent variable y^*_{ij} is unbounded and continuous, and it may be positive indicating a preference for B over A, or negative indicating a preference for A over B. It may also be 0 which would indicate indifference between the two choices. However the analyst does not observe this latent variable but instead observes the outcome, y_{ij}—a binary outcome equal to 1 if Option B is chosen and 0 if A is chosen. Thus;

$$y_{ij} = \begin{cases} 1 & \text{if } y^*_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$

The participants' response to each question was therefore included in the model as the binary dependent variable.

The independent variables were the differences in the levels of the variables. The independent variables were quasi-dummy coded for all levels of each dimension.

For example, if the health state 1 (the "best" health state) appeared on Option A but not B a code of −1 was coded. If it appeared in both A and B, or neither A or B, 0 was coded; if it appeared in Option B but not A, 1 was coded. This was performed for each level of each dimension and ensures that all coefficients and significance tests are of interest. This coding ensures that comparisons are made between attributes and their levels (including the comparison between adjacent attribute level estimates), rather than limited to some arbitrary omitted category. This coding ensures that the effects are uncorrelated with the constant term in a manner similar to that shown by Bech et al in a fixed comparison DCE [53]. "Best" in this instance refers to a priori expectations that higher levels of health status and confidence would be preferred to lower levels, more frequent visits from friends/relatives would be preferred to fewer and that quicker GP access would be preferred to slower. Table 1 shows each attribute and its levels. The omitted level in each case was level 1 (only health problem is moderate pain/discomfort, totally confident, can see GP tomorrow and see friends/relatives daily). The results described below and in Table 3, show decrements in utility associated with movements from level 1 of the attribute to levels 2 and 3.

As there were many observations from single responders the assumption of independence of the error terms is questionable. The standard response to this in health economics literature appears to be implementation of a random effects probit model [17,27], and this was considered in the first instance. However, random effects models account for correlation by allocating each individual their own choice-invariant constant term. Thus

Table 2 Characteristics of nonresponders and responders by questionnaire type

	Non-responders (N = 144)	Responders (N = 367) (average)	Responders (N = 188) (short quest)	Responders (N = 179) (long quest)
Characteristic				
Age	52.5	57.5	56.9	58.0
% female	72.9	68.7	73.9	63.1
Accommodation status:				
Owner occupied	57.6	74.4	72.9	76.0
Rented from LA	31.9	18.0	18.1	17.9
Rented privately	8.3	6.0	6.4	5.6
Other	2.1	1.6	2.7	0.6
Condition:				
Musculoskeletal	31.9	37.6	39.9	35.2
Endocrine	13.2	11.2	10.1	12.3
Circulatory	8.3	6.5	4.8	8.4
ME	6.3	7.4	6.9	7.8
Other	40.3	37.3	38.3	36.3
Employment status:				
Employed	16.0	20.4	18.1	22.9
Retired	33.3	39.8	36.7	43.0
Unable to work	36.8	31.6	37.8	25.1
Unemployed	6.3	3.5	4.3	2.8
Other	7.6	4.6	3.2	6.1

LA, local authority; ME, myalgic encephalopathy.

random effect models simply allow for heterogeneous preferences for choosing B over A all other things being equal. Thus it is a rather curious solution to a problem as it is inconsistent with the underlying theory and hence it is arguably more of a specification test than a genuine solution [54]. A more sophisticated model to allow for different respondents having different (heterogeneous) preferences is suggested. Heterogeneous preferences are examined using latent class models [55] in a subsequent analysis. This method allows for distinct groups (classes) within the population to have different preferences for particular outcomes or attributes in the experiment. The objective of the latent class regression model is to not only estimate these preferences but also to estimate the proportion of these classes within the population and assign a probability of membership for each individual.

The inclusion of constant terms is a violation of the theoretical basis of the model, but can be used as a notional misspecification test, in that if the constant term is significant there is some evidence of model misspecification. There is little guidance in the health economics literature on the procedures for dealing with a significant constant; indeed DCEs in the health economics literature appear to suppress the constant and assume it to be insignificant or use individual-specific (choice invariant) constant terms via random effects to "account" for the clustered nature of multiple responses from each individual [30,56].

Sensitivity Analyses

A number of sensitivity analyses were conducted. While the majority of the health economics literature has excluded "irrational" responses, this has been questioned by some authors [47,51,52] who suggest that there may be a number of alternative explanations for these responses. Therefore, a standard random effects probit was employed to test whether the inclusion of these responses altered the results or conclusions.

Two questionnaires were administered. It is possible that responses to the two questionnaires were systematically different. Thus a standard random effects probit was used on both questionnaires to identify any potential differences.

While it is argued above that a significant constant has no meaning in this experiment, the model could be estimated without a constant (i.e., the constant term is suppressed). This analysis was conducted as a sensitivity analysis.

GLLAMM techniques were used to identify potential heterogeneity of responses.

The appropriateness of each of the model specifications was assessed using receiver operating characteristic (ROC) curve. These curves allow examination of the ability of alternative models to discriminate between events and nonevents, in this case responses to the questionnaire.

Results

Five hundred eleven patients who had participated in the RCT of the EPP were sent a postal questionnaire. Of these 367 (71.8%) completed one of the questionnaires. Responders were, on average, slightly younger, more likely to own their own home and be in paid employment compared with nonresponders. The characteristics of patients who responded compared with nonresponders to the questionnaire are presented in Table 2. The characteristics of those returning both types of questionnaire are also presented in Table 2.

Questionnaire

For each attribute, each level was presented an equal number of times in both questionnaires. Thus in the short questionnaire, each level was presented three times (excluding the levels in the repeated question). In the long questionnaire, each level was presented nine times; thus the condition of level balance was satisfied. Using the fold over technique implies that there is no overlap, that is, there were no instances where an attribute was presented at the same level in both options of a single scenario. Orthogonality was ensured by the questionnaire design.

Discrete Choice Experiment

Patients whose responses were inconsistent (in that they gave different answers to the repeated question) were excluded from the primary analysis (N = 98, 26.7%). The remaining 269 responses were considered in the primary analysis. Although it is commonplace
to exclude these "irrational" responses, there is some discussion over this [47,51,52]. Therefore, these responses were included as a sensitivity analysis.

Table 3 shows the coefficients of each attribute or outcome. The coefficients represent the impact of a unit increase of each attribute on the probability of getting one answer (with all the other variables at their mean), and thus the relative importance of movements between levels of each attribute can be estimated by dividing one coefficient by another.

For example, if the coefficient around health state 2 is 0.5 (as health state 1 is the omitted category, this represents a move from

Table 3 Valuations of outcomes with inconsistent responses excluded

Attribute	Coefficient	Standard error
Health level 2	−0.081	0.033
Health 3	−0.464	0.057
Self-efficacy 2	−0.107	0.024
Self-efficacy 3	−0.483	0.038
General Practitioner access 2	−0.055	0.018
General Practitioner access 3	−0.284	0.032
Isolation 2	0.024	0.021
Isolation 3	−0.426	0.040
Constant	−0.148	0.017

Health 3/health 2 = 5.728; Self-efficacy 3/health 3 = 1.041.

health state 1 to health state 2) and the coefficient on self-efficacy level 2 is 0.25 (similarly, a movement from self-efficacy level 1 to self-efficacy level 2), responders valued the movement from health state 1 to health state 2 twice as much as the movement from self-efficacy level 1 to self-efficacy level 2.

The coefficients reflect the reduction in utility associated with moving from one state to another and are intuitively appealing in that the coefficients move in the anticipated direction. The only exception is the movement from isolation level 1 (“friends and relatives visit daily”) to isolation level 2 (“friends and relatives visit every few days”), where the latter is preferred. This movement is not statistically significant and is also plausible in that many individuals with chronic illness would find daily visits burdensome.

As expected, the movement from level 1 to level 3 is also greater than movement from level 1 to 2 for all attributes.

The coefficients around health status, self-efficacy and, to a lesser extent isolation, are of similar magnitude. In particular, the results indicate that patients value a movement from health state 1 (moderate pain, but no problems on other dimensions) to health state 3 (moderate pain, moderate anxiety/depression, some problems with self-care and some problems with mobility, no problems with usual activities) as approximately the same as a movement from self-efficacy level 1 (“totally confident in ability to manage condition”) to self-efficacy level 3 (“not at all confident in ability to manage condition”). A difference in “utility” between health state 1 and 3 of 0.2507 can be generated from the EQ5D tariff [57]. Thus the movement from not at all confident to totally confident would equate to a gain of 0.25 QALYs if maintained for 1 year (and assuming that we can ascribe utility values from patient generated responses rather than those of the general public).

However, it is worth noting that the constant term included in each of these models is statistically significant. This constant has no substantive meaning in an unlabelled experiment. In essence, this result implies that responders prefer A to B, even accounting for differences in the scenarios presented to them. Indeed, the size of the coefficient around the constant leads to a more worrying conclusion. If scenarios A and B are identical except that B has health state at level 1 (the “best” health state) and scenario A has health state at level 2 (the moderate health state), A would still be chosen as the coefficient around the constant is greater than that of health state level 2 (the movement from health state level 1 to health state level 2).

A potential explanation for the significant constant would be the use of simplifying heuristics when presented with choices that were closely matched. For example, individuals would systematically prefer the right hand option (B) when presented with difficult choices. This was tested by running a probit model with random effects for each individual question giving each question a specific constant term, a preference for B over A, after allowing for the differences between B and A. However, this analysis did not show any consistent preference for A over B (or vice versa).

Sensitivity Analyses

The inclusion of “inconsistent” responses did not alter the results. The direction of the coefficients remained the same in all cases, although the magnitude of the coefficients was slightly lower with the inclusion of these data. Table 4 shows the results of this analysis.

Estimating the model without a constant was performed as a sensitivity analysis. The results of this analysis are presented in Table 5. Although the direction of two of the attribute levels changes, most remain the same and all movements from the “best” level to the “worst” level yield reductions in utility. It is noticeable, however, that the magnitude of the coefficients has reduced considerably.

Table 4 Valuations of outcomes with inconsistent responses included

Attribute	Coefficient	Standard error
Health level 2	−0.079	0.033
Health 3	−0.462	0.057
Self efficacy 2	−0.081	0.024
Self efficacy 3	−0.468	0.038
General Practitioner access 2	−0.072	0.019
General Practitioner access 3	−0.254	0.032
Isolation 2	0.001	0.021
Isolation 3	−0.462	0.040
Constant	−0.120	0.017

Health 3/health 2 = 5.848; Self-efficacy 3/health 3 = 1.013.

The use of GLLAMM models did not change the direction of the results for the majority of individuals, although these models identified a minority of responders exhibiting counter-intuitive preferences. Standard models used previously in the literature would not have identified these individuals. Although the majority of individuals have a higher probability of being in class 1 (78%) rather than class 2, it is noticeable that those in class 1 exhibits the same direction of values of attributes as the previous models (good health is preferred to bad and so on), although the size of these coefficients is larger than previously estimated. For this class of responder, the health 3 to health 2 ratio in this model was 3.285, while the confidence 3 to health 3 ratio was 0.793, indicating that the individuals in this class considered a drop in health state (from level 1 to 3) as considerably worse than a drop in confidence (from 1 to 3).

However, there are a sizable minority (approximately 22%) who have counterintuitive values particularly for health state. This latent class appears to value decrements in health states while still valuing GP access (the other two attributes appear much less important to this group).

Health state 3, which is the “worst” health state, is preferred to health states 1 and 2. Although this raises some doubts about whether the responders who may be classified as class 2 responders have understood the exercise, it is this class that does not have a significant constant term. This raises the question of what to do with the preferences of this second class—do they genuinely represent odd preferences or are they evidence that a subsection of responders has misunderstood the questionnaire?

In the event that it is the second reason, the effects of misresponders in the homogeneous model are to weight the preferences of the population toward 0. Note how the ratio of moving down health states and confidence in the ability to manage condition increase relative to GP access (the level which apparently unaffected by the latent class specification). Table 6 shows the results of this sensitivity analysis.

In order to examine the specification of each of the models included in the main analyses and the sensitivity analyses, ROC curves were examined for each model. There were no sizable differences between probit and logit models including and excluding the constant term. However, the GLLAMM model allowing for heterogenous preferences “improved” the performance of the model in terms of the ability to distinguish between answers.

Interaction Effects

The design of the long questionnaire permitted the estimation of interaction terms between attributes. The results of this analysis

Table 5 GLLAMM model

Variable	Latent class 1		Latent class 2	
	Coefficient	Standard errors	Coefficient	Standard errors
Health level 2	−0.288	0.039	0.511	0.071
Health level 3	−0.946	0.058	0.651	0.094
Confidence level 2	−0.243	0.038	0.160	0.063
Confidence level 3	−0.750	0.049	−0.022	0.067
General Practitioner Access level 2	−0.017	0.035	−0.174	0.058
General Practitioner Access level 3	−0.240	0.044	−0.440	0.072
Isolation level 2	0.015	0.036	0.119	0.064
Isolation level 3	−0.649	0.040	−0.063	0.098
Constant	−0.219	0.028	−0.025	0.045
P(class)	0.784		0.216	

Health 3/health 2 (class 1) = 3.285; Self-efficacy 3/health 3 (class 1) = 0.793; Health 3/health 2 (class 2) = 1.274; Self-efficacy 3/health 3 (class 2) = −0.034. P(class) refers to the probability of being in class 1 or class 2.

are presented in Table 7. Few of the interaction terms had a substantial impact on the results. Three of the "significant" results were interactions between GP access and health. It is not clear a priori which direction these should take, although it is perhaps surprising that these are all negative, thereby implying that speedy GP access even in relatively poor health states is not very important. Noticeably, the interaction between health level 3 and self-efficacy is positive, indicating that if patients are in a relatively poor health state, improving their level of confidence improves their level of utility.

Discussion

A response rate to a postal questionnaire of over 70% indicated that patients were willing to complete the DCE. Results were consistent with a priori expectations, adding to the theoretical validity of the model, although a substantial minority failed a consistency check.

Self-efficacy is an important outcome, and patients were willing to trade decrements in HRQoL for improvements in self-efficacy. This has implications for the interpretation of economic evaluation in health care. Typically, economic evaluations employ the QALY as a measure of HRQoL. However, QALYs generated from the EQ5D instrument do not include self-efficacy explicitly. This analysis implies that, as self-efficacy is important and can be valued in terms of HRQoL, interventions that improve self-efficacy are more likely to be cost-effective than they are currently credited with.

However, there are a number of caveats. First, it is not clear whether self-efficacy is a health outcome; if it is not then should it be included in the objective function of a decision-maker working within a budget constrained health system? Decision-makers need to be clear a priori about what they are trying to maximize and not use vague statements about other factors that may be taken into account.

Table 6 Valuations of outcomes with suppression of constant

Attribute	Coefficient	Standard error
Health level 2	−0.098	0.026
Health 3	−0.286	0.043
Self-efficacy 2	0.101	0.026
Self-efficacy 3	−0.235	0.036
General Practitioner access 2	0.021	0.020
General Practitioner access 3	−0.109	0.029
Isolation 2	0.126	0.023
Isolation 3	−0.167	0.039

Health 3/health 2 = 2.918; Self-efficacy 3/health 3 = 0.8217.

Second, this study surveyed people with chronic conditions who had previous experience of self-managing and it is likely that these valuations of self-efficacy will be different from those of the general public. Indeed, the EPP from which these patients were recruited incorporates self-efficacy as one of its aims. Thus, a limitation of this study is that these patients' values may not be generalizable and are likely to overstate (or perhaps even set a maximum value for) the importance of self-efficacy, and thus make interventions that improve self-efficacy appear more cost-effective.

Third, it is conceivable that changes in self-efficacy (or part of those changes) are already incorporated into the QALY, through one or more of the dimensions of the chosen instrument.

It is also acknowledged that the use of DCEs may not force responders to focus on the real opportunity cost sacrifice to health by presenting a direct trade-off between health outcome and self-efficacy. This is likely to result in a greater chance of over-stating the relative value of the self-efficacy. Cookson [31] demonstrated that individuals expressed larger relativities when trade-offs were expressed in monetary terms rather than lives saved—similarly in this instance trading off self-efficacy may be more palatable and therefore result in higher valuations than if individuals were asked to reduce, for example, life expectancy.

In addition, the attributes included in the study may not have been ideal. In general, it may be better to describe objective outcomes as attributes rather than ask responders to imagine an emotional or psychological outcome and ask then to value that state. In this instance, given the need for a quantification of the value of other important outcomes in this patient group, the design was considered to be appropriate, although it is acknowledged that there may be other methods to generate these estimates.

Table 7 Valuations of interaction effects

Interactions	Coefficient	Standard error
Health 1/self-efficacy 1	−0.041	0.035
Health 1/self-efficacy 2	0.034	0.028
Health 2/self-efficacy 3	−0.033	0.026
Health 3/self-efficacy 1*	0.050	0.029
Health 1/General Practitioner access 1†	−0.065	0.025
Health 2/General Practitioner access 1†	−0.099	0.034
Health 2/General Practitioner access 2†	−0.080	0.030
Health 3/General Practitioner access 3†	0.084	0.022
Self-efficacy 1/General Practitioner access 3	−0.005	0.030
Self-efficacy 2/General Practitioner access 2	0.000	0.022
Self-efficacy 3/General Practitioner access 1	0.017	0.026
Self-efficacy 3/General Practitioner access 3	0.015	0.033

*Significant at 10%.
†Significant at 5%.

Finally there are two methodological issues worthy of comment. While the DCE seems to yield plausible coefficients around attributes two particular issues are concerning. First, the continued identification of a significant constant term cannot be ignored and requires further investigation. In an unlabelled experiment such as the one described above (and many others in the health economics literature), any value other than 0 indicates a systematic preference for A over B ($\hat{\beta}_0 < 0$) or B over A ($\hat{\beta}_0 > 0$) all other things being equal. Where the coefficient on the constant term is greater than that of another attribute, the implication is that improvement in that attribute would be outweighed by the constant term. Evidence to the contrary exists in this experiment and suggests a potential for model misspecification. We have examined a number of potential solutions, including the inclusion of interaction effects and the use of simplifying heursitics, to solve this problem but cannot explain the continued existence of the significant constant. There is little discussion of this in the health economics literature, and where a significant constant is identified, the problem tends to be ignored.

Second, the preferences of a sizeable minority of responders require further analysis. It is feasible that these are genuine preferences, but perhaps it is more likely that a subsection of responders has misunderstood the questionnaire. Alternatively, the lack of an opt-out option may have forced individuals into choices where they were genuinely indifferent. Indifference between options has been described previously as an explanation for "irrational" responses [52]. Other causes identified as explanations for "irrational" responses include the respondent including additional information not presented as part of the experiment [38,52], contradictory preferences, and protest votes [52]. Whichever of these possibilities is true is an empirical matter, but may impact on the results and interpretation of future DCEs, although the inclusion or exclusion of these responses does not impact on the conclusions of this study.

Conclusion

Although self-efficacy has been identified as important to patients with chronic conditions who have been exposed to interventions designed to impact on levels of self-efficacy, it is likely that in different patient groups other outcomes would be valued and traded for HRQoL. These "important" outcomes should be identified before commencing the study and appropriate techniques should be used to ascertain the rate of substitution between these outcomes and HRQoL.

The lack of concern in the health economics literature surrounding counterintuitive preferences and the possibility of model misspecification is disturbing. While these methodological caveats should not preclude the use of DCEs in health economic evaluation, these issues require further examination and suggest that results of DCEs where these issues could be a factor should be interpreted cautiously.

Financial Support

Gerry Richardson is funded by a Researcher Development award from the UK Department of Health.

References

1 Department of Health. The Expert Patient: A New Approach to Chronic Disease Management for the 21st Century. London: Department of Health, 2001.
2 Department of Health. Self care—A Real Choice: Self Care Support—A Practical Option. London: Department of Health, 2005.
3 Lorig KR, Ritter P, Stewart AL, et al. Chronic disease self-management program: 2-year health status and health care utilization outcomes. Med Care 2001;39:1217–23.
4 Lorig KR, Sobel DS, Ritter PL, et al. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4:256–62.
5 Kennedy A, Reeves D, Bower P, et al. The effectiveness and cost effectiveness of a national lay led self care support programme for patients with long-term conditions: a pragmatic randomised controlled trial. J Epidemiol Community Health 2007;61:254–61.
6 Bandura A. Self-efficacy toward a unifying theory of behavioral change. Psychol Rev 1977;84:191–215.
7 Bandura A, Jeffery RW, Gajdos E. Generalizing change through participant modeling with self-directed mastery. Behav Res Ther 1975;13:141–52.
8 Bandura A, Adams NE, Beyer J. Cognitive processes mediating behavioral change. J Pers Soc Psychol 1977;35:125–39.
9 Luszczynska A, Schwarzer R. Social cognitive theory. In: Conner M, Norman P, eds. Predicting Health Behaviour: Research and Practice with Social Cogniton Models. Maidenhead: Open University Press, 2005.
10 Gravelle H, Richardson G, Weatherly H, Ritchie G. An Assessment of the Quality of Economic Evaluations of Self Management: Primary Care Team. York: Centre for Health Economics, University of York, 2003.
11 Institute N, Clinical E. Guide to the Methods of Technology Appraisal. N0515. London: NICE, 2004.
12 Ryan M. Discrete Choice Experiments in health care. BMJ 2004; 328:360–61.
13 Lancaster K. A new approach to consumer theory. J Polit Econ 1966;74:132–57.
14 Bergmann A, Hanley N, Wright R., Energy I. Valuing the Attributes of Renewable Energy Investments. 2004 European Association of Environmental and Resource Economists conference. Budapest: European Association of Environmental and Resource Economists, 2004.
15 Manski C. The structure of random utility models. Theory Decis 1977;8:229–54.
16 Ryan M, Hughes J. Using conjoint analysis to assess women's preferences for miscarriage management. Health Econ 1997;6: 261–73.
17 Sculpher M, Bryan S, Fry P, et al. Patients' preferences for the management of non-metastatic prostate cancer: Discrete Choice Experiment. BMJ 2004;328:382.
18 Fraenkel L, Bogardus ST, Jr, Wittink DR. Risk-attitude and patient treatment preferences. Lupus 2003;12:370–6.
19 Gerard K, Lattimer V. Preferences of patients for emergency services available during usual GP surgery hours: a Discrete Choice Experiment. Fam Pract 2005;22:28–36.
20 Gyrd-Hansen D. Cost-benefit analysis of mammography screening in Denmark based on discrete ranking data. Int J Technol Assess Health Care 2000;16:811–21.
21 Gyrd-Hansen D, Sogaard J. Analysing public preferences for cancer screening programmes. Health Econ 2001;10:617–34.
22 Hall J, Kenny P, King M, et al. Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. Health Econ 2002;11:457–65.
23 Hill AR, Aspinall P, Dhillon B, et al. Utilities for quality of life from time trade-off and conjoint analysis in age related macular degeneration. Invest Ophthalmol Vis Sci 2004;45:3080.
24 Hundley V, Ryan M, Graham W. Assessing women's preferences for intrapartum care. Birth 2001;28:254–63.
25 Johnson FR, Banzhaf MR, Desvousges WH. Willingness to pay for improved respiratory and cardiovascular health: a multiple-format, stated-preference approach. Health Econ 2000;9:295–317.
26 Mark TL, Swait J. Using stated preference and revealed preference modeling to evaluate prescribing decisions. Health Econ 2004;13:563–73.
27 McKenzie L, Cairns J, Osman L. Symptom-based outcome measures for asthma: the use of discrete choice methods to assess patient preferences. Health Policy 2001;57:193–204.

28 Payne K, Elliott R. Using Discrete Choice Experiments to value preferences for pharmacy services. Int J Pharm Pract 2005;13:9–20.
29 Petrou S, Henderson J. Preference-based approaches to measuring the benefits of perinatal care. Birth 2003;30:217–26.
30 Ryan M, Netten A, Skatun D, Smith P. Using Discrete Choice Experiments to estimate a preference-based measure of outcome—an application to social care for older people. J Health Econ 2006;25:927–44.
31 Cookson R. Incorporating psycho-social considerations into health valuations: an experimental study. J Health Econ 2000;19:369–401.
32 Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. BMJ 2000;320:1530–3.
33 Ryan M, Bate A, Eastmond CJ, Ludbrook A. Use of Discrete Choice Experiments to elicit preferences. Qual Health Care 2001;10(Suppl. 1):S55–60.
34 Kennedy A, Gately C, Rogers A. Process evaluation of the EPP Report II: examination of the implementation of the Expert Patients Programme within the structures and locality contexts of the NHS in England (PREPP study). Available from: http://www.npcrdc.ac.uk/ [Accessed June 22, 2005].
35 Richardson G, Gravelle H, Weatherly H, Ritchie G. Cost-effectiveness of interventions to support self-care: a systematic review. Int J Technol Assess Health Care 2005;21:423–32.
36 Viney R, Savage E, Louviere J. Empirical investigation of experimental design properties of Discrete Choice Experiments in health care. Health Econ 2005;14:349–62.
37 Gyrd-Hansen D. Willingness to pay for a QALY. Health Econ 2003;12:1049–60.
38 Cheraghi-Sohi S, Bower P, Mead N, et al. Making sense of patient priorities: applying discrete choice methods in primary care using "think aloud" technique. Fam Pract 2007;24:276–82.
39 Mallinson S. Listening to respondents: a qualitative assessment of the Short Form 36 Health Status Questionnaire. Soc Sci Med 2002;54:11–21.
40 Louviere JJ, Hensher DA, Swait JD. Stated Choice Methods. Analysis and Application. Cambridge: Cambridge University Press, 2000.
41 Hahn G, Shapiro S. A Catalogue and Computer Program for the Design and Analysis of Orthogonal Symmetric and Asymmetric Fractional Factorial Experiments. New York: General Electric Research and Development Centre, 1966.
42 Lancsar E, Savage E. Deriving Welfare Estimates in Discrete Choice Experiments with Multiple Choice Options. 3rd Workshop Advancing the Methodology of Discrete Choice Experiements in Health Economics; 2005 6–7 July; University of Las Palmas de Gran Canaria. ECOMAS Research Unit.
43 Hanley N, Ryan M, Wright R. Estimating the monetary value of health care: lessons from environmental economics. Health Econ 2003;12:3–16.
44 Kocur G, Adler T, Hyman W, Aunet B. Guide to Forecasting Travel Demand with Direct Utility Assessment. US Dept of Transportation, Urban Mass Transportation Administration, Report UMTA-NH-11-0001-82-1. Washington, DC: US Department of Transportation, 1982.
45 Carlsson F, Martinsson P. Design techniques for stated preference methods in health economics. Health Econ 2003;12:281–94.
46 Phillips KA, Maddala T, Reed Johnson F. Measuring preferences for health care interventions using conjoint analysis: an application to HIV testing. Health Serv Res 2002;37:1681–705.
47 Lancsar E, Louviere J. Deleting "irrational" responses from stated choice experiments: a case of investigating or imposing preferences? 3rd Workshop Advancing the Methodology of Discrete Choice Experiements in Health Economics; 2005 6–7 July; University of Las Palmas de Gran Canaria. ECOMAS Research Unit.
48 Lancsar E, Savage E. Deriving welfare measures from Discrete Choice Experiments: a response to Ryan and Santos Silva. Health Econ 2004;13:919–24.
49 Lancsar E, Savage E. Deriving welfare measures from Discrete Choice Experiments: inconsistency between current methods and random utility and welfare theory. Health Econ 2004;13:901–7.
50 Santos Silva JMC. Deriving welfare measures in Discrete Choice Experiments: a comment to Lancsar and Savage (2). Health Econ 2004;13:913–18.
51 San Miguel F. Testing the Assumptions of Completeness, Stability and Rationality of Preferences in Health Economics Using Discrete Choice Experiments [PhD]. Aberdeen: University: of Aberdeen, 2000.
52 San Miguel F, Ryan M, Amaya-Amaya M. "Irrational" stated preferences: a quantitative and qualitative investigation. Health Econ 2005;14:307–22.
53 Bech M, Gyrd-Hansen D. Effects coding in Discrete Choice Experiments. Health Econ 2005;14:1079–83.
54 Richardson G, Bojke C. Incorporating "psychological" outcomes into cost-effectiveness analysis. HESG paper, York 2006 2006.
55 Rabe-Hesketh S, Skrondal A. Generalized linear mixed effects models. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, eds. Advances in Longitudinal Data Analysis. Boca Raton, FL: Chapman & Hall Press, 2006.
56 McIntosh EMR. Using Discrete Choice Experiments to derive welfare estimates for the provision of elective surgery: implications of discontinuous preferences. J Econ Psychol 2002;23:367–82.
57 Dolan P, Gudex C, Kind P, Williams A. A Social Tariff for EuroQol: Results from a UK General Population Survey. York: Centre for Health Economics, 1995.